

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2002, 00:09:27 ; Search time 8498.8 seconds
(without alignments)
30.345 Million cell updates/sec

Title: US-09-851-670-16
Perfect score: 24
Sequence: 1 gtccaagcagcagcaattctgca 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 111874

Minimum DB seq length: 0
Maximum DB seq length: 60

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	15.8	65.8	51	13	A2793014 2M0045E17
2	15.2	63.3	59	13	A2361084 1M0104D08
3	14.4	60.0	54	13	A2921737 1006031H1
4	14	58.3	39	11	BF339449 602039103
5	13.8	57.5	39	11	BF342092 602012848
6	13.8	57.5	29	13	A2658555 1M0535113
7	13.8	57.5	39	10	A2658555 1M0535113
8	13.4	55.8	43	13	A2869134 2M0181E10
9	13.4	55.8	51	10	A2869134 2M0181E10
10	13.2	55.0	29	13	A2588457 1M0396L17
11	13.2	55.0	42	13	A2632796 1M0487E12
12	13.2	55.0	57	10	AL595919 AL595919

13	13	54.2	50	10	AU105510
14	13	54.2 <td>50</td> <td>10 <td>AU105518</td> </td>	50	10 <td>AU105518</td>	AU105518
15	13	54.2 <td>50</td> <td>10 <td>AU105525</td> </td>	50	10 <td>AU105525</td>	AU105525
16	13	54.2 <td>50</td> <td>10 <td>AU105526</td> </td>	50	10 <td>AU105526</td>	AU105526
17	13	54.2 <td>50</td> <td>10 <td>AU105530</td> </td>	50	10 <td>AU105530</td>	AU105530
18	13	54.2 <td>50</td> <td>10 <td>AU105531</td> </td>	50	10 <td>AU105531</td>	AU105531
19	13	54.2 <td>50</td> <td>10 <td>AU105533</td> </td>	50	10 <td>AU105533</td>	AU105533
20	13	54.2 <td>50</td> <td>10 <td>AU105534</td> </td>	50	10 <td>AU105534</td>	AU105534
21	13	54.2 <td>50</td> <td>10 <td>AU105535</td> </td>	50	10 <td>AU105535</td>	AU105535
22	13	54.2 <td>50</td> <td>10 <td>AU105536</td> </td>	50	10 <td>AU105536</td>	AU105536
23	13	54.2 <td>50</td> <td>10 <td>AU105537</td> </td>	50	10 <td>AU105537</td>	AU105537
24	13	54.2 <td>50</td> <td>10 <td>AU105539</td> </td>	50	10 <td>AU105539</td>	AU105539
25	13	54.2 <td>50</td> <td>10 <td>AU105540</td> </td>	50	10 <td>AU105540</td>	AU105540
26	13	54.2 <td>50</td> <td>10 <td>AU105541</td> </td>	50	10 <td>AU105541</td>	AU105541
27	13	54.2 <td>50</td> <td>10 <td>AU105544</td> </td>	50	10 <td>AU105544</td>	AU105544
28	13	54.2 <td>50</td> <td>10 <td>AU105547</td> </td>	50	10 <td>AU105547</td>	AU105547
29	13	54.2 <td>50</td> <td>10 <td>AU105550</td> </td>	50	10 <td>AU105550</td>	AU105550
30	13	54.2 <td>54</td> <td>13</td> <td>A2456719 1M0259B15</td>	54	13	A2456719 1M0259B15
31	13	54.2 <td>56</td> <td>13</td> <td>A2416196 1M0191P16</td>	56	13	A2416196 1M0191P16
32	12.8	53.3	27	13	A2829657 2M0107N13
33	12.8	53.3	35	2	HS0004217 Homo sapi
34	12.8	53.3	38	13	A2651087 1M0521P03
35	12.8	53.3	50	10 <td>AU102695</td>	AU102695
36	12.6	52.5	36	13	A2794093 2M0047012
37	12.6	52.5	43	13	A2830602 2M0109N24
38	12.6	52.5	54	10 <td>AU102882</td>	AU102882
39	12.6	52.5	50	13	A2841488 2M0139C02
40	12.6	52.5	55	10 <td>AA138442 mq88912.r</td>	AA138442 mq88912.r
41	12.6	52.5	55	13	A2958077 2M0225008
42	12.4	51.7	39	13	A2810534 2M0076C07
43	12.4	51.7	44	13	A2501395 1M0340B14
44	12.4	51.7	50	10 <td>AU102761</td>	AU102761
45	12.4	51.7	50	13	TA346C10P

ALIGNMENTS

RESULT 1
A2793014 51 bp DNA 16-FEB-2001
LOCUS 2M0045E17R Mouse 10kb plasmid UGCG1M library Mus musculus genomic
DEFINITION clone UGCG2M0045E17 R, DNA sequence.

ACCESSION A2793014.1 GI:12937364
VERSION A2793014.1
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL
COMMENT Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0045 row: E column: 17
Seq primer: CACACGAGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 51.
Location/Qualifiers
1. 51

FEATURES
source

/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="taxon:10090"
/clone="UUGC2M0045E17"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6j (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 15 a 12 c 13 g 11 t
ORIGIN

Query Match 65.8%; Score 15.8; DB 13; Length 51;
Best Local Similarity 89.5%; Pred. No. 5.9e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtcacagagcagacattc 19
Db 27 gtcacagcagcagcatt 45

RESULT 2
AZ361084 59 bp DNA GSS 02-OCT-2000
LOCUS 1M0104D08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION Clone UUGC1M0104D08 R, DNA sequence.
ACCESSION AZ361084
VERSION AZ361084.1 GI:10474784
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 59)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weis,R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0104 row: D column: 08
Seq primer: CACACGCAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 59.
Location/Qualifiers

FEATURES

source

1. 59
/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="taxon:10090"
/clone="UUGC1M0104D08"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6j (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 15 a 9 c 18 g 17 t
ORIGIN

Query Match 63.3%; Score 15.2; DB 13; Length 59;
Best Local Similarity 85.0%; Pred. No. 1.1e+04;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 aagcagagcagacattctgca 24
Db 19 AAGCAGAGTCAGTCTCGCA 38

RESULT 3
AZ921737 54 bp DNA GSS 20-MAR-2001
LOCUS 100603JH12.y1 1006 - Rescuedu Grid G Zea mays genomic, DNA
DEFINITION sequence.
ACCESSION AZ921737
VERSION AZ921737.1 GI:13393675
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 54)
Walbot,V.
Maize genomic sequences found using engineered Rescuedu transposon
Unpublished (2001)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1006031 row: H column: 12
Class: transposon-tagged.
Location/Qualifiers

FEATURES

1. 54
/organism="Zea mays"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"

```

/clone_lib="1006 - Rescuemu Grid G"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: leaf; Vector: Rescuemu (engineered from
pBluescript backbone); Site_1: BamHI; Site_2: BglII;
Rescuemu is a 4.9 kb, modified maize Mu transposon
designed to allow plasmid rescue from total genomic DNA.
Mu elements insert preferentially into transcription
units. For more information on Rescuemu, go to the web
site 'www.zmbl.iastate.edu' and follow the links for
'Rescuemu.' Grid G was grown at Stanford in 2000. DNA was
extracted from leaf punches, double digested using BamHI
and BglII, and ligated to form circular plasmids. DH10B
cells were transformed and then screened on LB plates with
ampicillin."
BASE COUNT      6 a      20 c      12 g      16 t
ORIGIN

Query Match
Best Local Similarity 60.0%; Score 14.4; DB 13; Length 54;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY      1 gccaagcagcagcaattctcgca 24
Db      38 GCGCCGAGGAGAGACTCTCTCCA 15

RESULT      4
LOCUS      BF339449/c      39 bp      mRNA      EST      22-NOV-2000
DEFINITION 602039103F1 NCI_CGAP_Brn64 Homo sapiens CDNA clone IMAGE:4186752
5', mRNA sequence.
ACCESSION  BF339449
VERSION     BF339449
KEYWORDS   GI:11285904
SOURCE     EST.
ORGANISM   human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 39)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLM9508 row: f column: 01
            High quality sequence stop: 38.
FEATURES
            source
            1..39
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4186752"
            /clone_lib="NCI_CGAP_Brn64"
            /tissue_type="gliblastoma with EGFR amplification"
            /lab_host="DH10B (T1 phage-resistant)"
            /note="Organ: brain; Vector: PCMV-SPOK6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 1.57 kb. Constructed by Life
            Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      6 a      11 c      13 g      9 t
ORIGIN

Query Match
Best Local Similarity 75.0%; Pred. No. 2.4e+04;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY      1 gccaagcagcagcaattctcgca 24
Db      38 GCGCCGAGGAGAGACTCTCTCCA 15

RESULT      4
LOCUS      BF339449/c      39 bp      mRNA      EST      22-NOV-2000
DEFINITION 602039103F1 NCI_CGAP_Brn64 Homo sapiens CDNA clone IMAGE:4186752
5', mRNA sequence.
ACCESSION  BF339449
VERSION     BF339449
KEYWORDS   GI:11285904
SOURCE     EST.
ORGANISM   human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 39)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLM9409 row: o column: 11
            High quality sequence stop: 37.
FEATURES
            source
            1..39
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:418962"
            /clone_lib="NCI_CGAP_Brn64"
            /tissue_type="gliblastoma with EGFR amplification"
            /lab_host="DH10B (T1 phage-resistant)"
            /note="Organ: brain; Vector: PCMV-SPOK6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 1.57 kb. Constructed by Life
            Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      6 a      11 c      13 g      9 t
ORIGIN

```

```

Best Local Similarity 77.3%; Pred. No. 3.5e+04;
Matches 17; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY      3 ccaagcagcagcaattctcgca 24
Db      38 CCAGGCGCTGAGCCACTTCTGCA 17

RESULT      5
LOCUS      BF342092/c      39 bp      mRNA      EST      22-NOV-2000
DEFINITION 602012848F1 NCI_CGAP_Brn64 Homo sapiens CDNA clone IMAGE:4148962
5', mRNA sequence.
ACCESSION  BF342092
VERSION     BF342092
KEYWORDS   GI:11288842
SOURCE     EST.
ORGANISM   human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 39)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLM9409 row: o column: 11
            High quality sequence stop: 37.
FEATURES
            source
            1..39
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4148962"
            /clone_lib="NCI_CGAP_Brn64"
            /tissue_type="gliblastoma with EGFR amplification"
            /lab_host="DH10B (T1 phage-resistant)"
            /note="Organ: brain; Vector: PCMV-SPOK6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 1.57 kb. Constructed by Life
            Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      6 a      11 c      13 g      9 t
ORIGIN

Query Match
Best Local Similarity 58.3%; Score 14; DB 11; Length 39;
Matches 17; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY      3 ccaagcagcagcaattctcgca 24
Db      38 CCAGGCGCTGAGCCACTTCTGCA 17

RESULT      6
LOCUS      AZ658555/c      29 bp      DNA      GSS      14-DEC-2000
DEFINITION IM0535113F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0535113 F. DNA sequence.
ACCESSION  AZ658555
VERSION     AZ658555
KEYWORDS   GI:11795617
SOURCE     GSS.
            house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 29)

```

Query Match

58.3%; Score 14; DB 11; Length 39;

AUTHORS
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmud,M., Meenen,E., Pedersen,T., Rellly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.

TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL
Unpublished (2000)

COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0535 row: I column: 13
Seq primer: CGTTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 29.

FEATURES
source
1. 29
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG1M0535113"
/clone_lib="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-gold, T1-resistant, F-"
/note="Vector: PMD42ny; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g11473214|9b|AF1297.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapored mouse DNA was annealed to
adapored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT
5 a 5 c 8 g 11 t

ORIGIN

Query Match 57.5%; Score 13.8; DB 13; Length 29;
Best Local Similarity 88.2%; Pred. No. 4.2e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 2 tccaagcagagcaatt 18
11111111111111111111

Db 29 TCCTACCCAGACCAATT 13

RESULT 7
A0011691/c

LOCUS
A0011691 39 bp mRNA EST 03-AUG-1998

DEFINITION
A0011691 Schizosaccharomyces pombe late log phase cDNA

ACCESSION
A0011691

VERSION
A0011691.1 GI:3356600

KEYWORDS
EST.

SOURCE
fission yeast.

ORGANISM
Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;

REFERENCE
1 (bases 1 to 39)
Moriyomo,M. and Mita,K.
Identification of expressed sequence tags of Schizosaccharomyces
pombe

JOURNAL
Unpublished (1998)

COMMENT
Contact: Mitsunori Moriyomo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: moriyom@nirs.go.jp.

FEATURES
source
1. 39
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc11857"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mp19; The cDNA library of
Schizosaccharomyces pombe was prepared by cloning cDNA
into the SmaI site of M13mp19 DNA and the direction of DNA
sequences was not always from 5' to 3'. The cDNA data of
schizosaccharomyces pombe are available for searching on
the World Wide Web. (URL: http://www.nirs.go.jp)"

BASE COUNT
10 a 11 c 5 g 12 t

ORIGIN

Query Match 57.5%; Score 13.8; DB 10; Length 39;
Best Local Similarity 83.3%; Pred. No. 4.3e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Y 7 ggcagagcaattctgca 24
11111111111111111111

Db 24 GNCAGAGTAATTCCTGGA 7

RESULT 8
A2869134/c

LOCUS
A2869134 43 bp DNA GSS 21-FEB-2001

DEFINITION
2M0181E10F Mouse 10kb plasmid UUCG1M library Mus musculus genomic
clone UUCG2M0181E10 F, DNA sequence.

ACCESSION
A2869134

VERSION
A2869134.1 GI:13073144

KEYWORDS
GSS.

SOURCE
house mouse.

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 43)

AUTHORS
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmud,M., Meenen,E., Pedersen,T., Rellly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.

TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL
Unpublished (2000)

COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0181 row: E column: 10
Seq primer: CGTTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 43.

FEATURES
source
1. 43
Location/Qualifiers

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C2M0181E10"
/clone_lib="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PMD42ny; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g11473211419b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT      8 a      17 c      5 g      13 t

ORIGIN

Query Match      55.8%; Score 13.4; DB 13; Length 43;
Best Local Similarity 93.3%; Pred No. 6.4e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      10 agagcaattctcga 24
        ||||| |||||
Db      41 AGAGCAAGTTCTGCA 27

RESULT 9
LOCUS      AU013478      51 bp      mRNA      EST      03-AUG-1998
DEFINITION AU013478 Schizosaccharomyces pombe late log phase cDNA
ACCESSION  AU013478
VERSION     AU013478.1 GI:3368269
KEYWORDS    EST.
SOURCE      fission yeast.
ORGANISM    Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
1 (bases 1 to 51)
Moriyomo, M. and Mita, K.
Identification of expressed sequence tags of Schizosaccharomyces
pombe
Unpublished (1998)
Contact: Mitsuoki Moriyomo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: moriyom@nirs.go.jp

FEATURES
Source
1..51
/organism="Schizosaccharomyces pombe"
/db_xref="taxon:4896"
/clone="spc08212"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/notes="Vector: M13mp19; The cDNA library of
Schizosaccharomyces pombe was prepared by cloning cDNA
into the SmaI site of M13mp19 DNA and the direction of DNA

```

```

sequences was not always from 5' to 3'. The cDNA data of
Schizosaccharomyces pombe are available for searching on
the World Wide Web. (URL: http://www.nirs.go.jp)"

BASE COUNT      18 a      8 c      8 g      17 t

ORIGIN

Query Match      55.8%; Score 13.4; DB 10; Length 51;
Best Local Similarity 73.9%; Pred. No. 6.5e+04;
Matches 17; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy      2 tccaagcagagcaattctga 24
        ||||| ||||| ||||| |||||
Db      14 TCCATGTATGATGTAATTTTCGCA 36

RESULT 10
LOCUS      A2588457/c
DEFINITION 1M0396L17R Mouse 10kb plasmid U06C1M library Mus musculus genomic
clone U06C1M0396L17 R. DNA sequence.
ACCESSION  A2588457
VERSION     A2588457.1 GI:11710647
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 29)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Relliy
M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0396 Row: L Column: 17
Seq primer: CACACAGAAACGATATGCC
Class: Plasmid ends
High quality sequence stop: 29.
Location/Qualifiers
1..29
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0396L17"
/clone_lib="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PMD42ny; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g11473211419b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and

```

purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 5 a 9 c 5 g 10 t

ORIGIN

Query Match 55.0%; Score 13.2; DB 13; Length 29;
Best Local Similarity 83.3%; Pred. No. 7.7e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 6 aggcagagcaattctgc 23
||||| ||||| |||
Db 29 AGGAGGGCAATTTCAGC 12

RESULT 11

AZ632796/c

LOCUS 42 bp DNA GSS 13-DEC-2000
DEFINITION 1M0487E12R Mouse 10kb plasmid UGCG1M library Mus musculus genomic
clone UGCG1M0487E12 R, DNA sequence.

ACCESSION

AZ632796

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse whole genome scaffolding with paired end reads from 10kb
Plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dduun@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0487 row: E column: 12
Seq primer: CACACAGCAACACGCTATGACC
Class: Plasmid ends
High quality sequence stop: 42.
Location/Qualifiers
1. 42
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCG1M0487E12"
/clone_lib="Mouse 10kb plasmid UGCG1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male); was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (911473211419b1AF12907.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent *E. coli* XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 8 a 5 c 8 g 21 t

ORIGIN

Query Match 55.0%; Score 13.2; DB 13; Length 42;
Best Local Similarity 83.3%; Pred. No. 7.8e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 tccaagcagagcaattc 19
||||| ||||| |||||
Db 28 TCCAGGCGCAGACATTT 11

RESULT 12

AL595919 57 bp mRNA EST 30-JUL-2001
LOCUS AL595919 XGC-gastrula silurana tropicalis cDNA clone Tcas004d19 5',
DEFINITION mRNA sequence.

ACCESSION

AL595919

KEYWORDS

EST.

SOURCE

western clawed frog.

ORGANISM

Silurana tropicalis

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Sanger Xenopus tropicalis EST project 2001
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: Tcas004d19.sp6
Sequencing primer: sp6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1. 57
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="Tcas004d19"
/clone_lib="XGC-gastrula"
/dev_stage="gastrula (stages 10-5-13 mixed)"
/lab_host="Escherichia coli DH10B"
/note="Vector: PCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into PCS107 with EcoRI at the 5' end and NotI at the 3'
end."

FEATURES

source

BASE COUNT 11 a 28 c 10 g 8 t

ORIGIN

Query Match 55.0%; Score 13.2; DB 10; Length 57;
Best Local Similarity 83.3%; Pred. No. 7.9e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 7 ggcagagcaattctgca 24
||||| ||||| |||||
Db 56 GGCAGGCGCATTCCTGCA 39

RESULT 13

A0105510

LOCUS	50 bp	mRNA	EST	05-Apr-2001
DEFINITION	A0105510 Sugano Homo sapiens CDNA library Homo sapiens CDNA clone			
ACCESSION	CAS008101, mRNA sequence.			
VERSION	A0105510			
KEYWORDS	A0105510.1 GI:13555031			
SOURCE	EST.			
ORGANISM	human.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (bases 1 to 50)			
AUTHORS	Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata			
	,H., Ota,T., Isegai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo			
	,K., Suyama,A. and Sugano,S.			
TITLE	Fine structural analysis of transcription start sites of human			
JOURNAL	mRNAs using full-length enriched and 5'-end enriched cDNA libraries			
COMMENT	Unpublished (2001)			
	Contact: Yutaka Suzuki			
	Department of Virology			
	Institute of Medical Science, University of Tokyo			
	4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan			
	Email: Y.suzuki@ims.u-tokyo.ac.jp			
	Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano			
	,S. Construction and characterization of a full length-enriched and			
	a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).			
FEATURES	Location/Qualifiers			
source	1..50			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="CAS008101"			
	/clone_lib="Sugano Homo sapiens CDNA library"			
BASE COUNT	7 a 14 c 18 g 11 t			
ORIGIN				

Query Match	54.2%	Score 13	DB 10	Length 50
Best Local Similarity	76.2%	Pred. No. 9.6e+04		
Matches 16	Conservative 0	Mismatches 5	Indels 0	Gaps 0
Oy	2	tccaagcagcagcaattcttgc	22	
Db	10	TCCGAGCGCGGGCTATTCTGC	30	
RESULT 14				
LOCUS	AI105518			
DEFINITION	AI105518 50 bp mRNA EST	05-APR-2001		
ACCESSION	AI105518 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone HRC01211, mRNA sequence.			
VERSION	AI105518			
KEYWORDS	AI105518.1 GI:13555039			
SOURCE	EST.			
ORGANISM	human.			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 50)			
TITLE	Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hatae,H., Ota,T., Isegai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo,K., Suyama,A. and Sugano,S.			
JOURNAL	Fine Structural analysis of transcription start sites of human mRNAs using full-length enriched and 5'-end enriched cDNA libraries			
COMMENT	Unpublished (2001)			
	Contact: Yutaka Suzuki			
	Department of Virology			
	Institute of Medical Science, University of Tokyo			
	4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan			
	Email: yusuzuki@ims.u-tokyo.ac.jp			
	Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,S.			
	Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).			
FEATURES	Location/Qualifiers			
source	1..50			

```

/organism="Homo sapiens"  

/db_xref="taxon:9606"  

/clone="HRC01211"  

/clone_lib="Sugano Homo sapiens cDNA library"  

BASE COUNT      5 a          16 g          13 t  

ORIGIN  
  

Query Match      54.2%; Score 13; DB 10; Length 50;  

Best Local Similarity 76.2%; Pred. No. 9.6e+04;  

Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  

Oy      2   tccaagcgcagcattctg 22  

         ||| | | | | | | | |  

Db      16   TCCGAGCGCGCCTATTTCG 36  
  

RESULT 15  

AUI05525 LOCUS      50 bp mRNA EST      05-APR-2001  

DEFINITION AUI05525 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  

            HRC03805, mRNA sequence.  

ACCESSION  AUI05525  

VERSION    AUI05525.1 GI:13555046  

KEYWORDS   EST.  

SOURCE     human.  

ORGANISM   Homo sapiens  

            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  

            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  

REFERENCE  1 (bases 1 to 50)  

AUTHORS   Suzuki,Y., Tsunoda,T., Talre,H., Mizushima-Sugano,J., Sese,J., Hata  

            ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  

            ,K., Suyama,A. and Sugano,S.  

TITLE     Fine Structural analysis of transcription start sites of human  

            mRNAs using full-length enriched and 5'-end enriched cDNA libraries  

JOURNAL   Unpublished (2001)  

COMMENT   Contact: Yutaka Suzuki  

            Department of Virology  

            Institute of Medical Science, University of Tokyo  

            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  

            Email: yusuzuki@ims.u-tokyo.ac.jp  

            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,  

            S. Construction and characterization of a full length-enriched and  

            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).  

FEATURES  

Source  

1..50  

    /organism="Homo sapiens"  

    /db_xref="taxon:9606"  

    /clone="HRC03805"  

    /clone_lib="Sugano Homo sapiens cDNA library"  

BASE COUNT      6 a          15 g          14 t  

ORIGIN

```

Search completed: March 9, 2002, 00:09:28
Job time: 11044 sec

FEATURES	Location/Qualifiers
source	1. .50

THIS PAGE BLANK (USPTO)
